ANOVASimulator.R

Administrator

Thu Mar 31 14:24:57 2016

```r
# ANOVA data simulator
# 31 March 2016
# NJG

# Data from Rizzari et al. Oikos 123:829-836
# Figure 2b for 6 treatments

# treatment labels
treatLab <- LETTERS[1:6]

# treatment means estimated from figure 2
treatMeans <- c(61,60,23,5,6,4)

# treatment sd estimated from figure 2
# se = sd/sqrt(n)
# sd = se*sqrt(n)
# n = 10
# for first 3 groups se =10, so sd = 31
# for second 3 groups se = 2, so sd = 6

treatSD <- c(31,31,31,6,6,6)
n <- 10 # sample size per treatment

# create vector of treatment labels
Treatments <- rep(treatLab,each=10)
Treatments <- factor(Treatments)

# simulate data set with parameters from paper
AlgalMass <- rnorm(n=60,mean=rep(treatMeans,each=10),sd=rep(treatSD,each=10))

summary(AlgalMass)
```
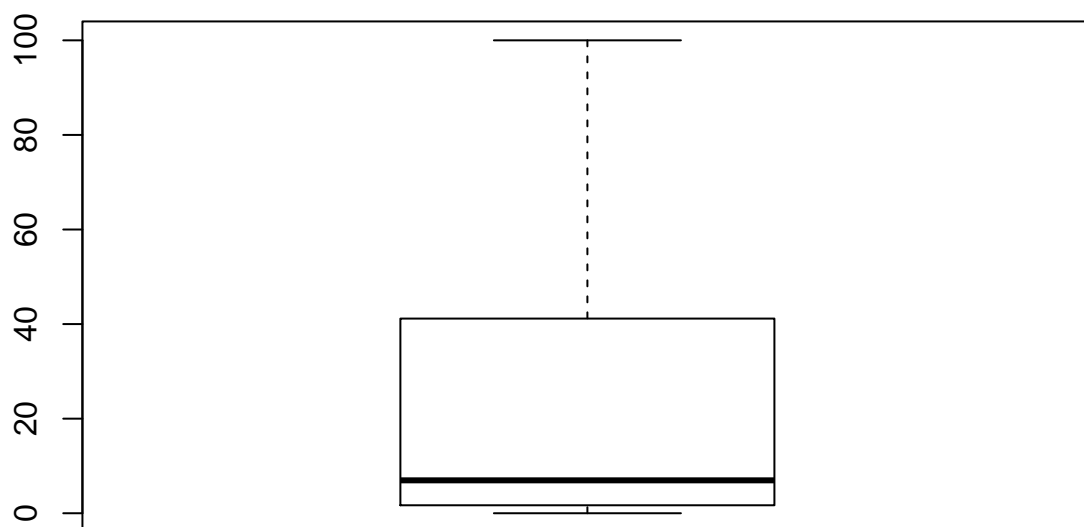
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -24.450   1.689   6.955  23.480  41.010 116.200
```

```r
#fudging the zeroes
AlgalMass[AlgalMass<0] <- 0

# fudging the upper end
AlgalMass[AlgalMass>100] <- 100
summary(AlgalMass)
```
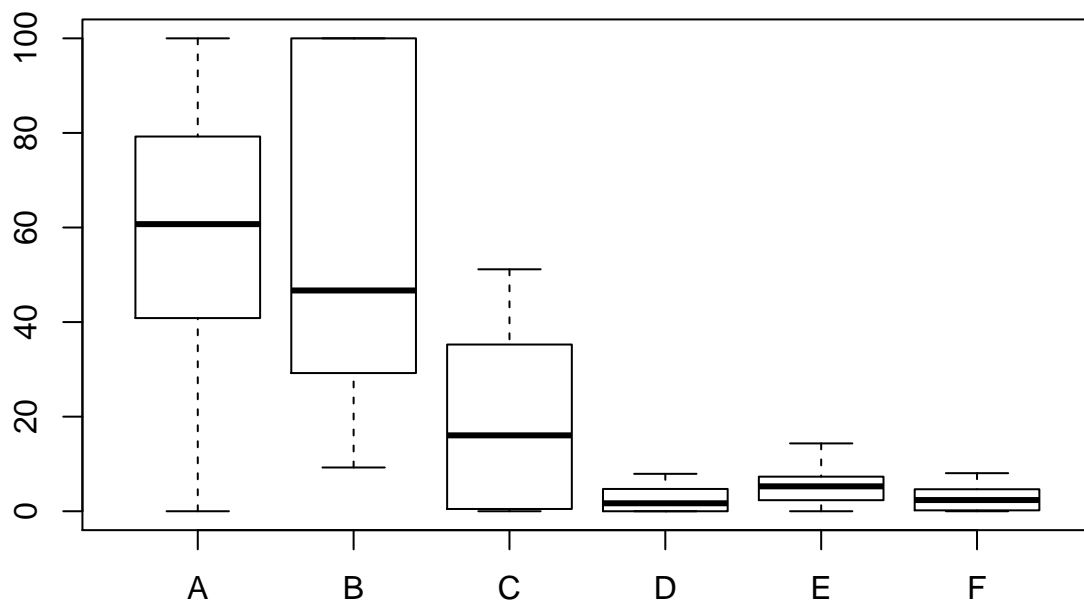
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.689   6.955  24.360  41.010 100.000
```

```
# show data as a boxplot
SimData <- data.frame(Treatments,AlgalMass)
boxplot(AlgalMass)
```



```
boxplot(AlgalMass~Treatments,data=SimData)
```

```r
# get group summaries; from Cookbook for R
library(plyr)
summaryData <- ddply(SimData,"Treatments", summarise,
                    N = length(AlgalMass),
                    mean = mean(AlgalMass),
                    median = median(AlgalMass),
                    sd = sd(AlgalMass),
                    se = sd /sqrt(N))
summaryData
```

```
##   Treatments N      mean    median        sd         se
## 1          A 10 55.878109 60.712897 30.882316  9.7658457
## 2          B 10 59.342935 46.681897 36.090242 11.4127365
## 3          C 10 20.147751 16.046307 19.145457  6.0543251
## 4          D 10  2.656255  1.681450  2.809019  0.8882898
## 5          E 10  5.405507  5.266479  4.335479  1.3709988
## 6          F 10  2.700652  2.366321  2.643684  0.8360061
```

```r
# analyze data as an ANOVA
MyModel <- aov(AlgalMass~Treatments,data=SimData)
z <- summary(MyModel)
```

```r
# creating a function and repeating the work
```

```r
treatLab <- LETTERS[1:6]
```

```
treatMeans <- c(61,60,23,5,6,4)
treatSD <- c(31,31,31,6,6,6)
sampleSize <- 10 # sample size per treatment

ANOVASimulator <- function(ss=sampleSize,l=treatLab,m=treatMeans,sd=treatSD){

# create random data set
RandomData <- rnorm(n=ss*length(l),mean=rep(m,each=ss),sd=rep(sd,each=ss))

# set up treatments as a factor
Treatments <- rep(l,each=ss)
Treatments <- factor(Treatments)

# combine elements in a data frame
SimData <- data.frame(Treatments, RandomData)

# run analysis of variance
MyModel <- aov(RandomData~Treatments,data=SimData)

# extract p value of test; answer from Stack Overflow!
output <- summary(MyModel)[[1]][["Pr(>F)"]][[1]]

# return the p value as the single output
return(output)

}
```

```
ANOVASimulator()
```

```
## [1] 1.89661e-06
```

```
# run the function multiple times to create a vector of p values

PVec <- replicate(100,ANOVASimulator())

# calculation proportion of tests that reject with p < 0.05
mean(PVec<0.05)
```

```
## [1] 1
```

```
# visualize with a histogram of p values
hist(PVec,breaks=seq(0,1.0,by=0.05),col=c("tomato",rep("goldenrod",19)))
```

4

Histogram of PVec